

A humán genom meglepetései

Dr. Duda Ernő
SZTE ÁOK Biológiai Intézet

A huszadik század biológiai forradalmának méltó befejezése volt az emberi genom bázissorrendjének meghatározása. Napjainkban már több száz élőlény (sok baktérium, egysejtű gombák, férgek, rovarok, a rizs, az egér és a csimpánz), örökítő anyagának ismert a szekvenciája, mégsem meglepő, hogy változatlanul az emberi genom áll az érdeklődés középpontjában.

Amikor az emberi genomról beszélünk, joggal merülhet fel a kérdés, hogy ugyan kinek a genomjáról van szó, hiszen minden ember más. Az egyik nagy meglepetés pontosan itt ért bennünket: kiderült, hogy hihetetlenül egyformák vagyunk. A különböző rasszok génjei között nincs nagyobb különbség, mint két átlag magyar ember génjei között. Az egyes egyének genetikai állománya pedig olyan mértékben hasonló, hogy tízezer „betűből” (nukleotidból) két ember között csak 10-12 eltérés fordul elő! Ez annak a bizonyítéka, hogy az ember nagyon fiatal faj és a jelenlegi sok milliárdos populáció egy igen kicsi közösségből, igen gyorsan szaporodott fel.

A kilencvenes években az élettudományok területén dolgozó tudósok egyértelmű véleménye volt, hogy az ember komplexitása csak nagyon nagy számú génnel kódolható. A becslések 70 ezertől a negyedmillióig terjedően szórtak, – ekkor már lehetett tudni, hogy egyes baktériumok 5-7 ezer génnel rendelkezhetnek, a legegyszerűbb férgek genomja pedig közel hússzezer gént tartalmaz. (Érdemes megemlíteni, hogy a hetvenes években viszont voltak olyan kutatók is, aki kémiai-matematikai számítások alapján úgy becsülték, hogy ha a gének szabályozása DNS-fehérje kölcsönhatáson alapul, akkor a genom nem rendelkezhet tíz-

harmincezerntől több (szabályozott kifejeződésű) génnél többel. Ezek a közlemények azonban a tudományos vélekedés periferáján maradtak.)

Ezek után igazi meglepetést – és nem csekély csalódást – keltett, hogy génjeink száma valahol 25 és 40 ezek között lehet, valószínűleg közelebb az alsó értékhez. Az eredményt hitetlenkedve fogadták világszerte. Alig lennének gén-gazdagabbak, mint egy egyszerű féreg? Hogyan lehet magyarázni az ember vitathatatlan magasabbrendűségét, élettani bonyolultságát egy olyan génkészlettel, amely méretében egyezik a pár centiméteres gyomnövényével (lúdfű) és töredéke a rizs genomjának (1. ábra)?

Azt még tudtuk magyarázni, hogy egyes növények DNS-e sokkalta hosszabb, mint az emberé, hiszen a növények genomja gyakran poliploid, számos búza variánsban 6 kromoszóma szerelvénnyel található a normális kettő helyett, ráadásul a nem-kódoló „kacat” szekvenciák (*junk DNA*) aránya is magas.

Itt azonnal egy további meglepetés is ért bennünket: kiderült, hogy az ember örökítő anyagának csak 1-2 százaléka (!) kódol fehérjét ill. eddig is ismert RNS-eket, pl. riboszómális és transzfer RNS-eket. A többi kacat! Az emberi genom elképesztően „szellős”, alig van benne „lényeg”. Amíg a kóli baktériumban 1000 bázispáronként (1 kb) találunk egy gént, az emberi genomban 1 millió bázispárnyi DNS (1 Mb) mindössze 10 gént kódol (3. ábra). Természetesen a mitokondrium DNS-e ugyanolyan „tömött”, mint a bakteriális DNS-ek.

Ennek a különbségnek a megértéséhez meg kell vizsgálnunk, mi a szerepe annak a DNS-nek, amely nem kódol fehérjét. A bakteriális és az állati gének szervezettségét összehasonlítva lényeges különbségeket láthatunk (2. ábra). A bakteriális gének szorosan egymás mellett helyezkednek el, csak rövid elválasztó szakaszok vannak közöttük. Az anyagcsere szempontjából összetartozó gének operonokba rendeződnek, amelyek aktivitását egyetlen szabályozó szekvencia határozza meg: vagy valamennyien aktívak vagy mind hallgat.

A gerincesek génjei ezzel szemben egyenként szabályozottak, gyakran bonyolultan összetett, hosszú kontroll szekvenciák hatása alatt állnak. A gének kódoló szakaszai nem folyamatosak: a fehérjék *működés szempontjából önálló feladattal rendelkező részeit* (a doméneket) kódoló DNS szakaszok (az exonokat) hosszú szekvenciák választják el egymástól. Ez utóbbiak (az intronok) információ tartalma már nem lesz része a mRNS-

eknek. Ahogyan szerszámainknak különféle alkatrészei vannak (pl. a kalapács feje és nyele), a fehérjék is működési egységekből állnak össze. Az emberi génekben gyakran találkozunk olyan jelenséggel, hogy az egyik alkotórész több másik alkotórésszel kombinálódhat. Mintha egy kalapács nyélhez többféle fejet is kódolna az illető gén. Ezután, a különféle szövetekben, sejtípusokban a helyi igényeknek megfelelően, egyetlen gén részeiből különböző célra alkalmas szerszámokat lehet előállítani. Ez a kombinálhatóság a Lego játékok és a korszerű, sokcélú szerszámok elvéhez hasonlít (3. és 4. ábra).

Az exonok viszonylag szabadon variálódhatnak, sokszorozódhatnak, ami nagyon felgyorsítja az evolúciót: ha egyszer egy jól használható domén kialakul, az számos fehérjében tud hasonló, de (partner doménjeitől függően) kissé más szerepet betölteni. A domének kicserélődését, kombinálódását segítik a beékelődött intron szakaszok.

Az elmúlt évek során fény derült arra, hogy valójában nem esett csorba „magasabbrendűségünkön”, csak a komplexitás titkát máshol kell keresnünk. Amíg a kb. ezer sejtből álló primitív féreg egy-egy génje egy-egy fehérje termék szintézisét kódolja, az emberben egyetlen gén fehérjék tucatjainak szintéziséhez elegendő genetikai információt tartalmaz. A különböző szövetek eltérő sejtípusaiban ugyanaz a gén más és domének kombinációiból álló fehérjék szintézisét kódolja (4. ábra). Ráadásul, a fehérjék bonyolult érési folyamata sejtípusoként más és más módon zajlik le, más tulajdonságú termékeket eredményezve.

Összegezve: ugyan nekünk sincs sokkal több génünk, mint az egyszerűbb szervezeteknek, de sokkal többféle géntermékkel (enzimmal, fehérjével) rendelkezünk, mint azok. Ha a *géntermékek* számát próbáljuk felbecsülni, már közel járunk a negyedmilliós számhoz, amit korábban génjeinkkel kapcsolatban feltételezték.

A következő nagy meglepetés az volt, hogy a 20–35 ezer emberi génnel szemben (ami a genom 1–2%-át teszi ki) DNS-ünk kb. felét(!) feltehetően vírus eredetű, „önző gének” milliói teszik ki (5. ábra). Ennek megértéséhez rövid kitérőre van szükség az un. „ugráló gének”, tudományos elnevezéssel a *mozgékony genetikai elemek* világába.

Vannak olyan DNS darabok (*transzpozonok*), amelyek *két végén* jellemző nukleotid sorrendű, ismétlődő szekvenciák találhatók, ezek között pedig egyetlen gén foglal helyet. A gén egy enzimet, a *transzpozázt* kódolja,

amely felismeri a két láncvégi szekvenciát, ki tudja vágni a transzpozont a genomból, majd be tudja építeni (*integrálni*) egy másik helyre. A transzpozon átugrik a genom egyik helyéről egy másik helyre. Ezáltal kárt okozhat (ha pl. egy génbe integrálódva tönkre teszi azt), de előnyös hatása is lehet, ha pl. exonokat is szállítva új gének kialakulását segíti elő.

Nagyobb jelentőségűek a *retro-transzpozonok* vagy *retroelemek*. Ezek közé tartoznak a retrovírusok is, mint pl. a közismert leukémia és szarkómavírusok, vagy a HIV is. Ezek az elemek (vírusok) nem transzpozált kódolnak, hanem *reverz transzkriptázt*, amely az RNS-ek információ tartalmát képes kétszálú DNS-sé átírni (fordítva, mint az RNS polimeráz, ezért *retro-* meg *reverz-*), majd a keletkezett DNS-t integrálni a genomba. A nagyobb jelentőség abból következik, hogy az így beépült retroelem (vagy vírus) génként viselkedik, RNS-ek készülnek róla, amiket a reverz transzkriptáz ismét visszaír DNS-ekké, majd ezek is beépülnek a genomba. Magyarán, egyre több lesz belőlük, szaporodnak a genomban!

Szerencsére, a reverz transzkriptáz nem pontosan (szépen, ahogy a csillag megy az égen) dolgozik, hanem rengeteg hibát vét, így az utód elemek jó része elveszti aktivitását (6. ábra). De még így is sokáig érthetetlen volt, hogyan tudja a genom megállítani a retroelemek mértéktelen szaporodását. Ui. kiderült, hogy a genomunkban található vírus-szerű elemek túlnyomó része a genetikailag inaktív heterokromatin régiókban található. Itt kanyarodunk vissza a humán genom nagy megtelepedéséhez.

Miután számos állat, növény és az élesztőgomba genom szekvenciája is rendelkezésre állt, össze lehetett hasonlítani azokat, hogy pl. megkeressük a legkonzerváltabb szakaszokat. Könnyű megérteni, hogy ha valaminek a szekvenciája az élesztőgombában és az emberben hasonló, annak alapvetően fontos feladata kell, hogy legyen. Rendre meg is találták a riboszómális fehérjék és RNS-ek, a hisztonok, a hősokk fehérjék és más, nagyon fontos alkotó elemek génjeit a legkevésbé változó szekvenciák között. A meglepetés akkor jött, amikor kiderült, hogy az ilyen roppant konzervált szekvenciák jeletős része nem *génekben* fordul elő, hanem az értéktelen, értelmetlen, blabla szekvenciának hitt „kacat” DNS-ben!

Mint a mesében, addig „kukáztak” a kutatók, míg kincsre leltek. Ezek a szekvenciák azonban sokkal rövidebbek, mint a gének, vagy akár az exonok, nem is kódolnak fehérjét (ezért nem találták meg ezeket korábban), hanem csak rövidke RNS-eket (*mikroRNS*, *siRNS*). Meglepően sokféle feladatot képesek ellátni ezek az RNS-ek. Egyik legfontosabb

feladatuk az, hogy képesek megvédeni bennünket külső és belső ellenségeinkkel szemben. Ha szekvenciájuk komplementer pl. egy vírus RNS-ével (vagy mRNS-ével), akkor egy rövid kétszálú szakasz alakul ki, amit felismer egy fehérje komplex, a RISC (*RNA-induced silencing complex*), ami azután darabokra vagdosza a virális RNS-t. Ha az illető RNS-t egy retroelem kódolta, akkor ezzel még nem ér véget a miRNS tevékenysége: a RISC további komplexet tud képezni a kromatin metilációját végző fehérjékkel (ezek hallgattatják el a nem kívánatos géneket). A RISC-hez kapcsolódott miRNS szekvencia-specifikusan irányítja a metilációt, ami a retroelem heterokromatinná alakításához vezet. Tehát a piciny RNS-ek aktivitása nem szorítkozik a mRNS pusztítására, további (vírus vagy retroelem) mRNS-ek szintézisét is képes megakadályozni, egyszer s mindenkorra.

A miRNS-eknek (7. ábra) a fenti védelmi feladat mellett egyéb szerepe is van. Egyes miRNS-ek hiánya fejlődési rendellenességeket okoz. Egyre több kísérlet látszik bizonyítani, hogy a miRNS-ek (és a sok tekintetben hasonló „silencing” siRNS-ek) részt vesznek saját génjeink működésének szabályozásában is. Egyrészt, a fentiekhez hasonlóan megakadályozhatják egyes gének működését, másrészt pozitív és negatív szabályozó feladatot is elláthatnak egy másik mechanizmus révén.

Sok miRNS szekvenciája *nem pontosan* komplementer a szabályozandó mRNS-ével, hanem csak többé-kevésbé. Ráadásul, nem a mRNS kódoló szekvenciájával képes dupla szálát alkotni, hanem a mRNS vége felé elhelyezkedő, át nem íródó, szabályozó szakasszal (3' *UTR*, *untranslated region*). Ebben az esetben is fehérje komplexek alakulnak ki, azonban ezek a komplexek nem darabolják fel a mRNS-t, hanem stabilitását és aktivitását szabályozzák. Mint az utolsó ábrán láthatjuk, a miRNS-ek génjei a fehérjét kódoló gének között helyezkednek el, gyakran csoportosan, vagy közös prekuzort kódolva.

Egyes hipotézisek szerint az emberi egyedek közötti különbséget nem is a gének közötti árnyalatnyi eltérések okozzák, hanem a „kacat” DNS-ben fellépő különbségek. Ez még csak hipotézis, de érdemes megemlíteni, hogy amíg a csimpánz és az ember génjei között alig két százalékos különbség van, addig pl. az *Alu* *szekvencia* néven ismert (retroelem eredetű) ismétlődő szakasz emberben 5–6-szor gyakoribb, mint majmokban.

Természetesen, ez nem jelenti azt, hogy az Alu miatt vált az ember emberré, de biztos, hogy az ilyen „kacat” szekvenciák, az ugráló gének meghatározó szerepet játszottak az evolúcióban. A genom véletlenszerű átrendezésével variabilitást teremtettek, majd a természetes szelekció elvégezte a legjobbak kiválogatást. Tehát a retrovírusok származékai és a kacat DNS-ek alakították ki az emberi genom mai szerkezetét. A genomunk felét kitevő vírus-maradványok pedig az ember genetikai multjának velünk élő (de nem működő) emlékei.

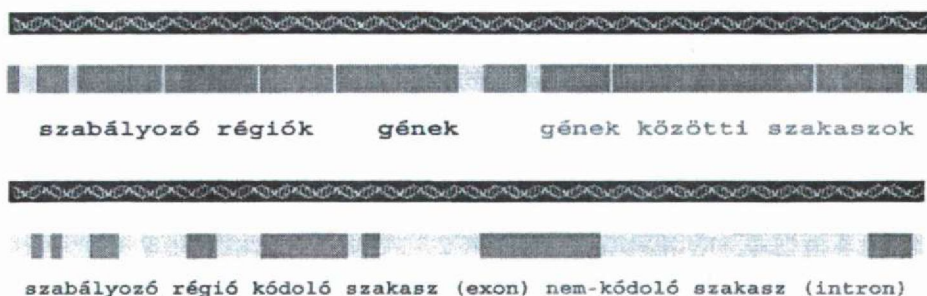
Egyes állatokban az ugráló gének nagy gyakorisággal változtatják a helyüket, folyamatosan alakítva a faj genetikai állományát. Ebből a szempontból az emberi faj meglehetősen stabil: szigorú kontroll alatt tartjuk hajdani mozgékony elemeinket. Egyedül az Y kromoszómában találunk még aktív elemeket, amelyek férfiakban okozhatnak örökletes károsodást (mint tudjuk, az emberi faj szempontjából a férfiak az „eldobható” kategóriába tartoznak, töredékük is elég lenne a faj fennmaradása szempontjából). Fejlődési rendellenességek, daganatos betegségek kialakulásában azonban szerepet játszanak a véletlenül aktiválódott mobilis elemek, egy-egy sejtben, sejtcsoportban felboríthatják a sejt szabályos működését.

A humán genommal való ismerkedésünk még távolról sem ért véget, génjeink többségének feladatáról még fogalmunk sincs. Lehet, hogy a kacatban más értékek is bújkálnak, bizonyára lesz még néhány meglepetésben részünk. Visszatekintve elképesztő, hogy az elmúlt néhány év mennyi alapvetően új felfedezéssel szolgált, amelyek teljesen átformálták elképzeléseinket saját DNS-ünkről és génjeinkről.

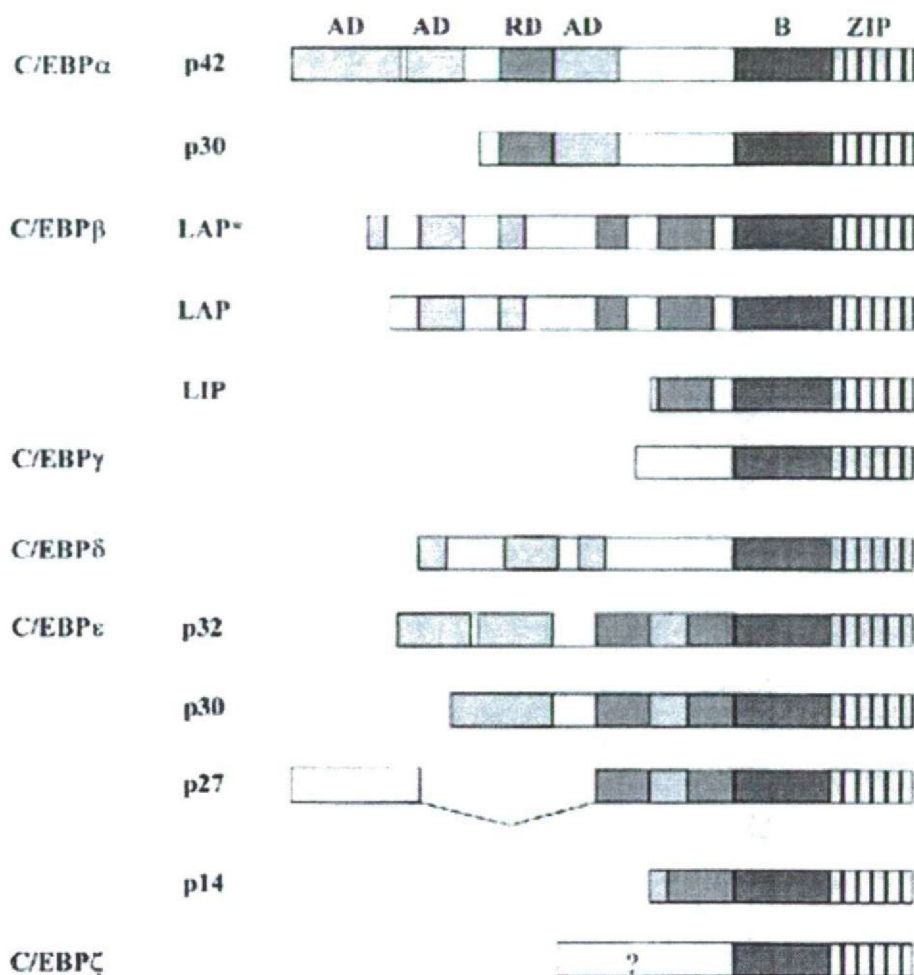
Organism	Haploid Genome Size (Mb)	Number of Genes	Genes per Mb
<i>Hemophilus influenzae</i> (bacterium)	1.8	1,700	940
<i>Escherichia coli</i> (bacterium)	4.6	4,400	950
<i>Saccharomyces cerevisiae</i> (yeast)	12	5,800	480
<i>Caenorhabditis elegans</i> (nematode)	97	19,000	200
<i>Arabidopsis thaliana</i> (plant)	118	25,500	215
<i>Drosophila melanogaster</i> (fruit fly)	180	13,700	76
<i>Oryza sativa</i> (rice)	430	60,000	140
<i>Danio rerio</i> (zebrafish)	1,700	22,000	13
<i>Mus musculus</i> (house mouse)	2,600	25,000	11
<i>Homo sapiens</i> (human)	2,900	25,000	10
<i>Fritillaria assyriaca</i> (plant)	120,000	ND	ND

1. ábra. Egyes élőlények genetikai állományának összehasonlítása

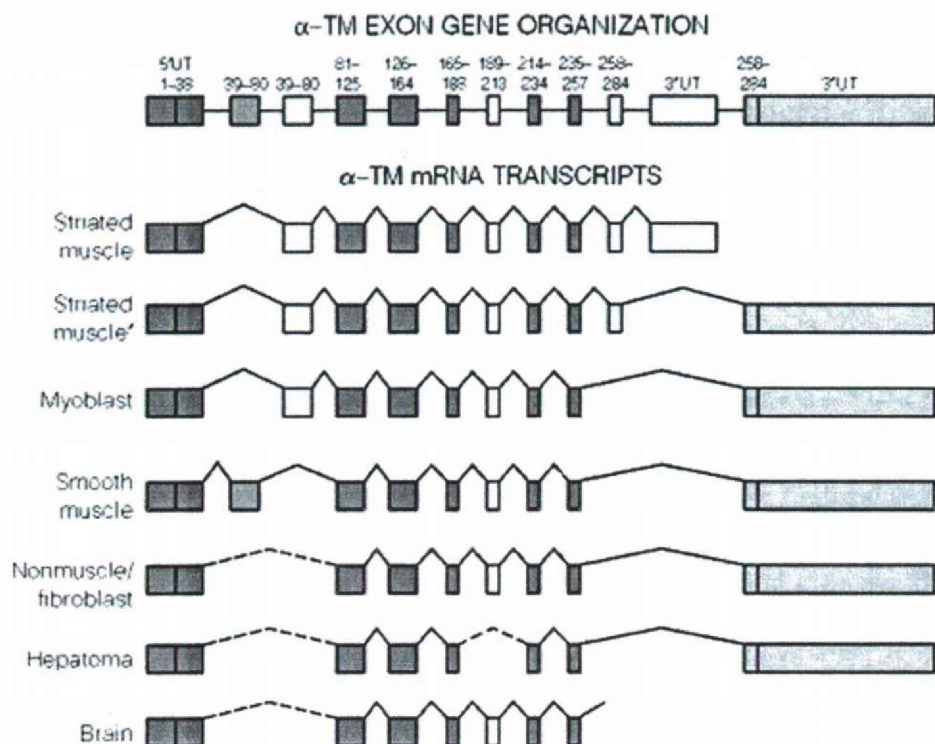
A **haploid genom** egy kromoszóma szerelvény nukleotid párjainak összegét adja meg (Mb, azaz millió bázispár mértékben). A második oszlopban a becsült gének számát látjuk, míg a harmadik oszlop a gén-sűrűséget adja meg: hány gén helyezkedik el egy 1 millió bázispár hosszúságú DNS darabon. A rizs 60 ezer génjével, vagy egy lilium a maga 120 milliárd bázispárnyi DNS-ével jóval meghaladja az emberi genom méretét.



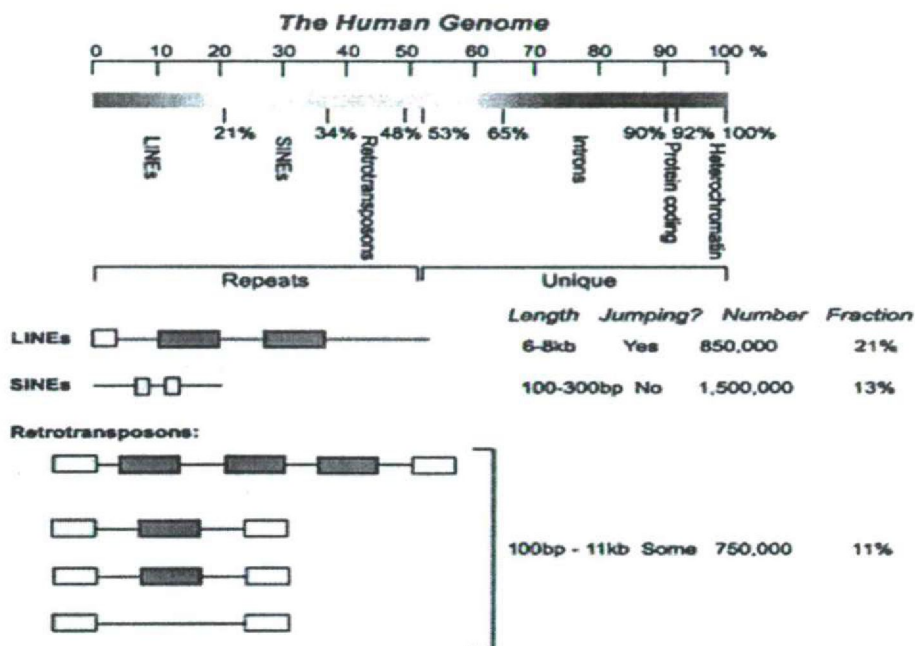
2. ábra. A prokarióta és az eukarióta genom szerkezete. A bakteriális gének sokkal sűrűbben helyezkednek el, mint pl. az állatok nem-kódoló szakaszokkal és hosszú gén közötti szekvenciákkal szétszabdalt exonjai (ezek a fehérjét kódoló szekvenciák). Az exonok azonban a magasabbrendűek genomjában szabadabban kombinálódhatnak, így az evolúció során könnyebben alakulhatnak ki új, hasznos gének.



3. ábra. Példa az exonok evolúciós jelentőségére. Az ábrán különféle, szteroid hatást közvetítő, génaktivitást szabályozó fehérjék szerkezete látható. Az azonos színű fehérje domének valószínűleg közös ősből származnak. A fehérje szerkezetében jól megfigyelhető az elemekből való építkezés és a szabad kombinálódás elve.

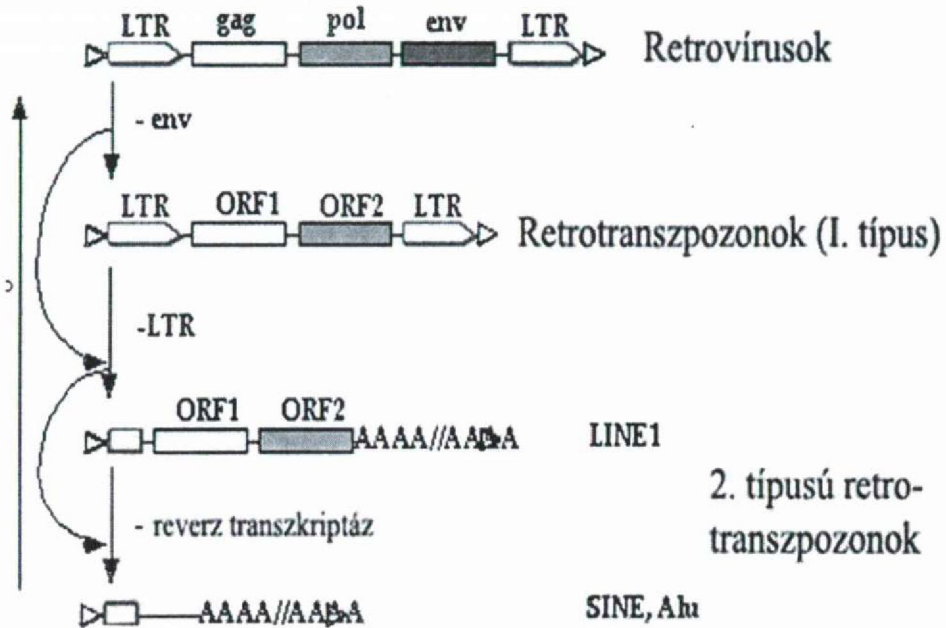


4. ábra. Egy gén, sokféle géntermék. Az ábrán azt látjuk, hogy a tropomiozin gén 13 exonból (fehérjét kódoló szakaszból) áll. A különféle szövetekben – a sejtek eltérő igényének megfelelően – más és más elemekből álló tropomiozin fehérje termelődik. Ezek hasonló, de nem azonos működésűek, hasonló, de nem azonos, sejt-specifikus feladatot látnak el.

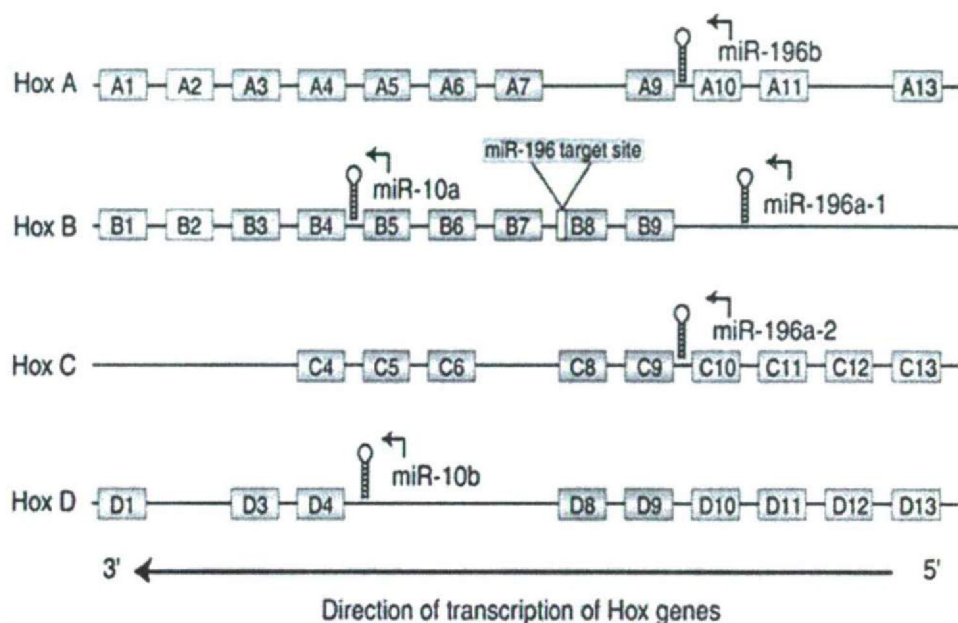


5. ábra. Az emberi genom felépítése. A genom nagyobbik része repetitív, ismétlődő szakaszokból (piros-sárga-zöld) tevődik össze, ezek többsége ugráló gén és retro-transzpozon, vagy retrovírus eredetű. Az egyedi szekvenciák (kék-lila) túlnyomó többsége fehérjét nem kódoló, gének közötti ill. exonok közötti (intron) szekvenciákból áll. A gének közötti és az intron szekvenciák gyakran tartalmaznak génaktivitást szabályozó elemeket

Retroelemek



6. ábra. A retrovírus eredetű szekvenciák családja. Az aktív retrovírusok génjeik elvesztése során I. típusú elemekké, majd egyre több szekvencia elvesztésével hosszú (LINE) ill. rövid (SINE) 2. típusú elemekké válnak. Ezek közül az ún. Alu szekvenciák az emberi genomban fordulnak elő leggyakrabban, a mi DNS-ünkre jellemzőek. Lehet, de nem bizonyított, hogy az emberi faj kialakulásában is szerepük lehetett.



7. ábra. A mikroRNS-eket kódoló DNS szakaszok a gének közötti „kacat” DNS-ben található. Az ábrán az egyedfejlődésben meghatározó szerepet játszó homeo gének (HoxA-D) exonjai közé ékelődött, szabályozó mikroRNS-eket kódoló DNS szakaszok (hajtűszerű jelek) láthatók. E szakaszok kiejtése súlyos torzfejlődést, differenciálódási rendellenességeket okoz.